PATENT
5760-16500
VRTS 0508

"EXPRESS MAIL" MAILING LABELNUMBER
EV 318248502 US
DATE OF DEPOSIT DECEMBER 15, 2003
I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C.F.R. § 1.10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO: COMMISSIONER FOR PATENTS, P.O. BOX 1450, ALEXANDRIA, VA 22313-1450.

Derrick Brown

Server-Free Archival of Backup Data

By:

Ernest Patrick Hanavan, III

Atty. Dkt. No.: 5760-16500/ VRTS 0508

Robert C. Kowert/JDC
Meyertons, Hood, Kivlin, Kowert & Goetzel, P.C.
P.O. Box 398
Austin, Texas 78767-0398
Phone: (512) 853-8800

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] This invention is related to the field of storage management and, more particularly, to combining server free and disk-based backup of data in storage networks.

Description of the Related Art

[0002] The need for data storage capacity in enterprise systems is growing exponentially with no end in sight. One common storage architecture is direct attached storage (DAS). In most DAS configurations, a user's data resides on storage media directly connected to a host/server predominantly using small computer system interconnect (SCSI) technology, and normally there is but a single (i.e. non-fault tolerant) path for retrieving that data. Organizations are beginning to recognize that the restrictions imposed by DAS SCSI architecture are too costly to continue as a viable solution. Such restrictions include the following:

- SCSI disk arrays must be located no more than 25 meters from the host server;
- The parallel SCSI bus is susceptible to data errors resulting from slight timing discrepancies or improper port termination; and
- SCSI array servicing frequently requires downtime for every disk in the array.

[0003] One solution has been to create technology that enables storage arrays to reside directly on the network, where disk accesses may be made directly rather than through the server's SCSI connection. This network-attached storage (NAS) model eliminates SCSI's restrictive cable distance, signal timing, and termination requirements. However, it adds a significant load to the network, which frequently is already starved for bandwidth. Gigabit Ethernet technology only alleviates this bottleneck for the short term, so a more elegant solution is desirable.

[0004] Another primary concern of large enterprises is having their data available continuously. In order to insure a high degree of user data availability, data storage should be highly reliable. High reliability data storage is normally implemented by maintaining a backup copy of the data on a physically independent storage device. Under the DAS architecture maintaining a backup copy of user data requires copying the data from storage attached to one host/server to storage attached to another host/server over the LAN used by client/users. As the amount of user data maintained by an enterprise increases, the amount of LAN bandwidth required to perform timely data backup increases proportionately. At some point the copying of backup data may significantly interfere with the client/user's ability to use the network. The storage area network (SAN) architecture may significantly reduce the impact of data backup on the primary network.

[0005] The SAN model places storage on its own dedicated network, removing data storage from both the server-to-disk SCSI bus and the main user network. This dedicated network most commonly uses Fibre Channel technology, a versatile, high-speed transport. The SAN includes one or more hosts that provide a point of interface with LAN users, as well as (in the case of large SANs) one or more fabric switches, SAN hubs and other devices to accommodate a large number of storage devices. The hardware (e.g. fabric switches, hubs, bridges, routers, cables, etc.) that connects workstations and servers to storage devices in a SAN is referred to as a "fabric." The SAN fabric may enable server-to-storage device connectivity through Fibre Channel switching technology to a wide range of servers and storage devices.

[0006] Since host primary data is stored on storage devices connected to the SAN, at least some of the SAN storage devices may use disk type technology to provide the access times required by the hosts. Traditionally, the cost of disk type storage has been relatively high and backup of primary data has been to the more mature and cost-effective tape type devices. Using this technology, primary data to be backed up may be read from

disk storage by a host/server running a backup application. The host may then write the data to a tape drive storage device. For relatively small amounts of primary data, this method may be sufficient and may not overtax the capabilities of the host or the tape drive.

[0007] However, as the amount of data to be backed up increases the difference in speed between tape and disk type storage becomes increasingly significant. Disk type storage devices can be faster than tape by an order of magnitude or more in backup mode and many orders of magnitude faster during restore operations due to its random access capability. Also, as disk drive technology matures, the cost may become competitive with that of tape type storage devices for the backup of primary data making this option financially more attractive.

[0008] One way to perform disk-based backup of primary data is analogous to the method described above for tape-base backup. A host/server may read primary data from disk storage and write a backup copy of the primary data to physically separate disk drive devices. Another method for disk-based backup may be referred to as server-free. In a server-free backup environment data is copied directly from a SAN attached storage manager/client disk storing primary data, to a SAN attached backup drive via a SAN data mover using third party copy (3PC) including the SCSI extended copy command. The host/server sends commands to the data mover to tell it which blocks to copy from which SAN attached disk to which other SAN attached device. This method considerably reduces the overhead for primary data backup on the host/server.

[0009] Disk-based backup of primary data alone may not be sufficient to meet all the backup requirements of an enterprise. Specifically, data archival and off-site storage for disaster recovery may still require backup using a highly stable, mature technology such as magnetic tape. Also, some large enterprises are reluctant to adopt server-free technology as their method of primary data backup since it does not have an extensively proven track record.

SUMMARY

[0010] Embodiments of a system and method for making an archival copy of data by performing a third party copy of backup data across a SAN are described. A SAN may include a SAN fabric, multiple host/servers and multiple storage devices including archival storage devices. The SAN fabric may include the switches, hubs, bridges, routers, etc. necessary to allow any host/server to access any storage device. The host/servers allow applications access to primary data stored in the SAN. At least one of the host/servers may run a backup server that is responsible for making backup and archival copies of the primary data stored on the storage devices by the applications. The backup server may initiate a server-free copy of the backup data through the SAN fabric to the archival storage devices.

[0011] A server-free copy may also be referred to as a "Third Party Copy" (3PC). This SAN backup technique may use the SCSI Extended Copy command to copy data from one set of storage devices to another. These can be disks, tapes, or other types of storage devices. This SCSI protocol command can be used on devices in the SAN fabric.

[0012] Internally, the third party copy device that receives and performs the Extended Copy command may be known as a "data mover". The data mover may be an intelligent device somewhere in the SAN infrastructure that understands the extended copy command. This can be another server, but more likely will be a fabric device, such as, a SAN switch or storage router, or a smart-storage device, such as an intelligent tape device or disk device. In order to perform a 3PC backup, the backup host/server may need to perform some preparatory tasks prior to invoking the data mover.

[0013] If the destination of the extended copy is a removable media device, e.g. a tape drive, then the host may issue other SCSI commands to get the removable device into the proper position (i.e. loading or positioning the tape). The data mover may require a frozen image of the data, which is to be backed up. The backup server may provide this

frozen image by any available method including, but not limited to snapshotting, cloning, and mirror splitting. The data mover may only be able to access the data to be backed up by physical disk addresses. The backup server may "map" the logical representation of the data to its absolute physical address ranges and provide this along with all the file or volume information necessary for the server-free copy command. Then the host may issue the server-free copy command to the data mover.

[0014] The server-free copy command (e.g. extended copy command) basically tells the third party copy device (data mover) to copy data from one storage device directly to another storage device across the SAN. After issuing the extended copy command, no further instructions have to be issued by the host to copy the data - the devices themselves perform the entire data copy operation through the SAN fabric under the supervision of the data mover. Upon completion of the extended copy command, the backup server may instruct the device storing the backup data to "thaw" the backup copy of the data. This may cause changes to the primary data collected during the archival process to be made to the backup copy. The backup server may also update the backup database to reflect the new status. By combining a traditional back up of primary data to disk with a server-free back up of the disk-based backup data, an additional level of data protection may be obtained and the impact on application servers may be minimized with server-free data copy, while allowing primary data to be backed up in any manner, such as a server-based backup to disk.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] The following detailed description makes reference to the accompanying drawings, which are now briefly described.

[0016] Figure 1 illustrates an exemplary SAN system capable of implementing embodiments of third party copying of backup data to archival storage devices.

[0017] Figure 2 illustrates the paths for command/data flow for a LAN-free backup, according to one embodiment.

[0018] Figures 3A and 3B illustrate respectively the command and data flow for a server-free or third party copy, according to one embodiment.

[0019] Figure 4 is a flowchart of a method for performing a server-free copy through a SAN fabric, according to one embodiment.

[0020] While the invention is described herein by way of example for several embodiments and illustrative drawings, those skilled in the art will recognize that the invention is not limited to the embodiments or drawings described. It should be understood, that the drawings and detailed description thereto are not intended to limit the invention to the particular form disclosed, but on the contrary, the intention is to cover all modifications, equivalents and alternatives falling within the spirit and scope of the present invention as defined by the appended claims. The headings used herein are for organizational purposes only and are not meant to be used to limit the scope of the description or the claims. As used throughout this application, the word "may" is used in a permissive sense (i.e., meaning having the potential to), rather than the mandatory sense (i.e., meaning must). Similarly, the words "include", "including", and "includes" mean including, but not limited to.

DETAILED DESCRIPTION OF EMBODIMENTS

[0021] Figure 1 illustrates an exemplary SAN system capable of implementing embodiments of server-free backup of backup data to archival storage devices, according to one embodiment. A SAN may be a high-speed network that enables fast, reliable access among host/servers 120 and storage resources 140. A SAN may be used to connect servers to storage devices, servers to each other, and storage devices to other storage devices through combinations of hubs 138, switches 132, bridges 134, and/or other devices 136 collectively referred to as a SAN Fabric 130.

[0022] A hub 138 may perform a repeater function, which is to amplify and/or regenerate a received signal and provide it at multiple outputs. A hub takes the data that comes into one port and sends it out all other ports that are included in the hub. It doesn't usually perform any filtering or redirection of data. Bridges 134 are useful for joining networks made of different media types or protocols together into larger networks, and keeping network segments free of data that doesn't belong in a particular segment. Bridges 134 may join networks or network segments, which use different protocols. In such cases, the bridge may perform lower layer translation on the packets that it transfers from one network to another.

[0023] A given SAN may include any number of the components described above depending on the number of host/servers and storage devices that it will interface. In some embodiments, the SAN may include only a subset of the types of SAN fabric components. SAN fabric 130 may also include a data mover 170. In some embodiments data mover 170 may be implemented as a stand-alone component separate from the other components of SAN fabric 130, but more frequently, the data mover may be implemented as a part of one of the other components of the SAN fabric. The data mover function may be located anywhere within the SAN fabric, within any of the components previously described. In addition the data mover 170 may be implemented as a component of the

data storage devices 140 or archival storage devices 160. A more detailed description of the function of the data mover 170 is presented below.

[0024] One or more end-user platforms (clients 115) may access the SAN, typically via a LAN or WAN connection to one or more of the hosts 120 to access primary data 180. Primary data 180 may be any data that an application running on a platform attached to network 110 requires in order to perform its function. To guard against the inability to access primary data due to a failure of a device storing the primary data, for example, backup data may be created by copying primary data to storage devices that are physically separate from those storing the primary data on some periodic or non-periodic basis.

[0025] Storage devices 140 and 160 may include one or more of, but are not limited to, RAID (Redundant Array of Independent Disks) systems, disk arrays, JBODs (Just a Bunch Of Disks, used to refer to disks that are not configured according to RAID), tape devices, and optical storage devices. Normally, an enterprise's primary data may be stored on disk drive type storage devices to provide fast access time. In order to provide fault tolerance for devices storing primary data 180 and maintain the same level of data availability, it may be desired to store a backup copy 185 of the primary data on a type of storage device which may have a similar access time to that of the disk type device storing primary data 180.

[0026] Creation of backup copies of primary data 180 may be the responsibility of a backup server 122 running on, for example, host/server 120A. Backup server 122 may maintain a backup database 175 that indicates what primary data is backed up and where the backup copies are located. Each host/server 120 is typically connected to the fabric 130 via one or more Host Bus Adapters (HBAs). SAN fabric 130 may enable server-to-storage device connectivity through Fibre Channel switching technology. SAN fabric 130 hardware may include one or more fabric components, as described above, as well as the interconnecting cables (for Fibre Channel SANs, fibre optic cables).

[0027] Host/servers 120 may be any of various types of devices, including, but not limited to, personal computer systems, desktop computers, laptop or notebook computers, mainframe computer systems, workstations, network appliances, network computers, Internet appliances, or other suitable devices. Host system 120A may include at least one processor. The processor may be coupled to memory. Memory is representative of various types of possible memory media, also referred to as "computer readable media." Hard disk storage, floppy disk storage, removable disk storage, flash memory and random access memory (RAM) are examples of memory media. The terms "memory" and "memory medium" may include an installation medium, e.g., a CD-ROM or floppy disk, a computer system memory such as DRAM, SRAM, EDO RAM, SDRAM, DDR SDRAM, Rambus RAM, etc., or a non-volatile memory such as a magnetic media, e.g., a hard drive or optical storage. The memory medium may include other types of memory as well, or combinations thereof.

[0028] SANs may be implemented using a wide variety of technologies. The SAN mix can include Enterprise Systems Connection (ESCON), Fiber Distributed Data Interface (FDDI), Asynchronous Transfer Mode (ATM), IBM's Serial Storage Architecture (SSA), and Fibre Channel. SAN architectures may also implement a number of underlying protocols, including TCP/IP and variants of SCSI (Small Computer System Interface). The most popular implementation of SAN for open systems is based on SCSI over Fibre channel. Fibre Channel Protocol (FCP) specifies how to run the SCSI command set over a dedicated Fibre Channel optical fabric. In direct server attached storage, a local SCSI controller on a peripheral bus fulfills a data request initiated by a SCSI driver in the host server. On a SAN, a Fibre Channel host bus adapter (HBA) may replace the SCSI controller in each server 120 to connect to the SAN fabric 130, which in turn may connect to disk arrays, tape drives, and other storage devices. A LUN (logical unit number) is the SCSI identifier of a logical unit within a target, the system component that receives a SCSI I/O command. A logical unit is an entity within a SCSI target that executes I/O commands. SCSI I/O commands are sent to a target and executed by a logical unit within that target. A SCSI physical disk typically has a single logical unit.

Tape devices and array controllers may incorporate multiple logical units to which I/O commands may be addressed. Each logical unit exported by an array controller may correspond to a virtual disk. SAN fabric 130 may be implemented, for example, in accordance with the Fibre Channel Switch Fabric-2 (FC-SW2) open standard specification to provide redundancy and ensure high data availability. SANs may be deployed in both homogeneous and heterogeneous environments. In a heterogeneous environment, a SAN may allow different kinds of servers, e.g. Windows NT, UNIX, Linux, Netware, and OS/390, to share different kinds of storage, e.g. disk, tape, and redundant arrays of inexpensive disks (RAID). With this shared capacity, organizations may be able to acquire, deploy, and use storage devices more cost-effectively.

[0029] As mentioned previously, backup data may be created from its primary data on a periodic or non-periodic basis to insure that the data is available in the event of failure of the storage device(s) storing the primary data or the accidental deletion of data by a user. In one embodiment, the creation of the backup data may be performed by a backup server copying primary data to a storage device(s) physically distinct from that storing the primary data. In other embodiments, the disk subsystem may set up one or more mirrors of the data. Every write request to the original data may be automatically duplicated to the other mirrors, or copies of that data, within that subsystem. When a copy of the data is required for a backup process, for example, the mirror may be temporarily suspended, or "split". The disk subsystem may be told to temporarily stop the mirrored writes going to one of the mirrors, and to save the writes in a separate location. A stable copy of the primary data may then be backed-up off the split mirror. When the backup is complete, mirroring may be resumed. The disk subsystem may first apply all the saved writes to the mirror from which the data was copied, and then resume normal mirror operations. The original file or volume may not be affected by splitting the mirror.

[0030] In addition to backing up primary data, an additional server free method of backup may be used to reduce the workload on the backup server, as discussed in detail below. In one embodiment, primary and backup data may be stored on disk type storage

devices for speed of backup and restore operations, and additionally a server-free back up of the backup data may be performed from the backup disk storage.

[0031] Further backup, or archival, data may be created from backup data to insure data availability in the event of a more catastrophic event causing the failure of storage devices storing primary and backup data. In one embodiment, backup data may be copied by data mover 170 from a disk type storage device(s) to a tape type drive in order to produce the archival data without having to pass the data through one of the host servers. In other embodiments, the backup data may be copied by data mover 170 to any type storage device that permits the archival copy of the data to be located remotely from the backup data. In order to prevent overburdening the backup server with the additional copying of backup data, a server-free copy technique may be used to produce the archival data. The server-free method of data copying is discussed in detail below with Figs. 3A & 3B.

[0032] Figure 2 illustrates the paths for command/data flow for a backup of primary data, according to one embodiment. The backup server 122 may read the primary data 180 from storage device on the SAN 140B. The primary data may move through the SAN along path 152. The backup server 122 may read the data to its internal memory and then write a backup copy 185 through path 154 to storage device 140D that is physically separated from storage device 140B. In this way, the backup copy 185 of the data may be available via the SAN in the event of the loss of the storage device 140B storing the primary copy of the data 180. As shown in Fig. 2, performing a backup of primary data may involve data moving through the SAN, through a host backup server, and back through the SAN again to a backup disk. In other embodiments, primary data may be backed up using any other suitable backup technology, such as a server-free backup.

[0033] In SAN architecture systems, it may be desirable to limit the involvement of resources attached to the primary network in data backup to the greatest extent possible. In these cases, the workload of the backup server may be considerably reduced by using

server-free backup. In a server-free backup, data may be copied via a data mover directly from the SAN attached storage device storing the data to be copied to another SAN attached storage device that is to store the copy data. The transfer of the data may actually be done by the data mover within the SAN fabric and therefore, the backup server may not have to read and write the data at all. The backup server may send commands to the data mover to tell it which blocks to copy from which SAN attached storage device to which other SAN attached storage device. The data may actually be copied rather than moved from one location to another. The data mover is a means for copying backup data through the SAN fabric to the one or more archival storage devices in response to the backup server initiating a server-free copy. Server-free copy provides a way to back up and restore large volumes of data between client-owned storage devices using a method that considerably reduces overhead on the backup server and the client.

[0034] In addition to backing up primary data to disk-based storage, the backup data itself may be backed up to an archival storage. Figures 3A and 3B show the command and data flow paths for performing a server-free copy of backup data 185 from storage 140D to an archival device 160B across the SAN, according to one embodiment. Referring to Figure 3A, when it is desired to produce an archival copy 190 of backup data 185, backup server 122 may issue commands 192 to archival storage device 160B to make preparations to receive the data. For example, if archival storage device 160B is a tape drive, commands 192 may cause the correct media to be mounted for storing archival data 190. Backup server 122 may provide source and destination information along with data extents to data mover 170 as indicated by 196.

[0035] Figure 3B illustrates the flow of data across the SAN during the server-free copy of backup data from backup storage to archival, according to one embodiment. During the server-free copy operation, the movement of the data may be under the control of data mover 170 and require no participation from a host/server 120. Referring once again to figure 3A, when the creation of the archival copy has been completed, backup server 122 may send updates 194 to modify the contents of backup database 175 in

storage device 140A to reflect the new status/location of the archival of backup data 185. Data that has been backed up using server-free data copy may be restored over a server-free path, or over the SAN through a host/server. By combining a traditional back up of primary data to disk with a server-free back up of the disk-based backup data, an additional level of data protection may be obtained and the impact on application servers may be minimized with server-free data copy, while allowing primary data to be backed up in any manner, such as a server-based backup to disk.

[0036] As mentioned previously, one way to perform server-less or server-free backup is the implementation of the SCSI-3 extended copy command. This is an extension to the original SCSI-3 specification. A "Third Party Copy" (3PC) SAN backup technique may use the SCSI Extended Copy command to copy data from one set of devices to another. These can be disks, tapes, or other types of storage devices. This SCSI protocol command can be used on devices connected via the SAN fabric.

[0037] A third party copy device that receives and performs the Extended Copy command may be referred to as the "data mover". The data mover is an intelligent device somewhere in the SAN infrastructure that understands the extended copy command. This may be another server, but more likely will be a fabric device, such as a SAN switch or router or a smart-storage device such as an intelligent tape device or disk device. In order to perform a 3PC backup, the backup host/server may need to perform some preparatory tasks prior to invoking the data mover, but does not need to read and write the data itself.

[0038] Figure 4 is a flowchart of a method for performing a server-free copy of backup data stored on disk across a SAN to an archival storage device, according to one embodiment. If the destination of the server-free copy is a removable media device, e.g. a tape drive, then the host may issue other SCSI commands to get the removable device into the proper position (e.g. loading or positioning the tape) as shown in block 400. The data mover may require a frozen version of the backup data, which is to be backed up. Thus, any further modifications to the backup data may temporarily be quiesced. The

backup server may provide this frozen image by any suitable method including, but not limited to snapshotting, cloning, and mirror splitting. The data mover may only be able to access the data to be backed up by physical disk addresses. The backup server may "map" the logical representation of the data to its absolute physical address ranges, as indicated at block 410, and provide this along with all the file or volume information necessary for the extended copy command. Then the host may issue the server-free copy command to the data mover.

[0039] The server-free (e.g. extended) copy command basically tells the third party copy device (e.g. the data mover) to copy data from one storage device directly to another storage device across the SAN. After issuing the server-free copy command, no further instructions have to be issued by the host to move the data for the copy - the devices themselves perform the entire data copy operation through the SAN fabric under the supervision of the data mover, as shown at 420. Upon completion of the extended copy command, the backup server may "thaw" the backup device, as indicated in block 430. This may cause changes to the primary data collected during the archival process to be made to the backup copy. The backup server may also update the backup database to reflect information regarding the copy operation (e.g., to indicate the existence of the additional copy, location, and/or other information), as indicated in block 440.

[0040] Various embodiments of the disclosed invention may further include receiving, sending or storing instructions and/or data implemented in accordance with the foregoing description upon a carrier medium. Generally speaking, a carrier medium may include storage media or memory media such as magnetic or optical media, e.g., disk or CD-ROM, volatile or non-volatile media such as RAM (e.g. SDRAM, DDR SDRAM, RDRAM, SRAM, etc.), ROM, etc. as well as transmission media or signals such as electrical, electromagnetic, or digital signals, conveyed via a communication medium such as network and/or a wireless link.

[0041] The various methods as illustrated in the Figures and described herein represent exemplary embodiments of methods. The methods may be implemented in software, hardware, or a combination thereof. The order of method may be changed, and various elements may be added, reordered, combined, omitted, modified, etc.

[0042] Various modifications and changes may be made as would be obvious to a person skilled in the art having the benefit of this disclosure. It is intended that the invention embrace all such modifications and changes and, accordingly, the above description to be regarded in an illustrative rather than a restrictive sense.